Scanning worm attack/defense experiments with PREDICT datasets

G. Kesidis, NSF/DHS EMIST Project, Penn State

DHS PREDICT Workshop, Newport Beach, CA, 09/27/05

- Defense/containment devices assumed deployed in peripheral enterprise network(s)
  - End-hosts and/or network nodes, e.g., access router
  - Stand alone or collaborative
- Need background traffic for evaluation of false-positives.
- Need attack traffic for evaluation of false-negatives.
- In practice, most defenses are evaluated using
  - worst-case traffic scenarios (→over-engineering), and
  - limited deployments in operational networks (representative?).
- Significant context-specific tuning required after deployment.

Trace with attack traffic naturally *in situ*

- Desirable to have Internet packet trace
  - At various physical locations that are potential deployment points for defenses under consideration
  - Background traffic without and with attack traffic
  - Several traces in same temporal context too for improved statistical confidence.
- Kind of tolerable anonymization depends on the defense (detection and response), e.g., detection of anomalously
  - large destination IP addresses contacted per unit time
  - large freq of failed scans, scans to dark addresses in particular
  - large number of packets with certain src/dst ports
  - few DNS precursors (may require DPI, i.e., payload info)
- Also, DPI suggested for detection of polymorphic worms *given* a signature of an instance of the malcode.
- Problem: such traces are unavailable and could only indicate performance for known attacks.
- Note: in EMIST, we do not model the host vulnerability nor the infection mechanism in detail.

Enterprise traffic with background and attack traffic artificially blended

- Well known examples exist that are now understood to be of limited value, obsolete.
- Need both intra-network and exogenous traffic sources.
- Detailed replaying background traffic difficult because, e.g.,
  - Significant protocol state missing from trace.
  - Attack traffic will alter background traffic, e.g., when attack traffic volume causes congestion.
- Again, what actually needs to be replayed depends on the defense under test.
- What about hypothetical worm propagation methods?
- Motivates need for modeling.

Modeling attack traffic that is exogenous to the enterprise network under test

- For Slammer and Witty worms, /8 tarpit traces of
  - scanning packets with unmodified source addresses and payloads removed.
  - associated routeviews.
- Given this information, can compute
  - total scan-rate
  - scan-rate per worm
  - number of worms per stub

Modeling attack traffic that is exogenous to the enterprise network under test

- Can recreate exogenous attack traffic using
  - Raw tarpit data (single-node)
  - Scaled-down emulation (64+ nodes)
  - Mathematical model (single-node)
- Can extend models to hypothetical scanning worms and past worms for which such data is unavailable.